

# Genome Sequence of the Tobacco Bacterial Wilt Pathogen *Ralstonia solanacearum*

Zefeng Li,<sup>1</sup> Sanling Wu,<sup>1</sup> Xuefei Bai,<sup>1</sup> Yun Liu,<sup>1</sup> Jianfei Lu,<sup>2</sup> Yong Liu,<sup>3</sup> Bingguang Xiao,<sup>3</sup>  
Xiuping Lu,<sup>3</sup> and Longjiang Fan<sup>1\*</sup>

Department of Agronomy & James D. Watson Institute of Genome Sciences, Zhejiang University, Hangzhou 310058, China<sup>1</sup>;  
Zhejiang Pesticide Testing Management, Fengqi Road, Hangzhou 310020, China<sup>2</sup>; and Yunnan Academy of  
Tobacco Agricultural Sciences and China Tobacco Breeding Research (Southern) Center, Yuxi 653100, China<sup>3</sup>

Received 17 August 2011/Accepted 19 August 2011

***Ralstonia solanacearum* is a causal agent of plant bacterial wilt with thousands of distinct strains in a heterogeneous species complex. Here we report the genome sequence of a phylotype IB strain, Y45, isolated from tobacco (*Nicotiana tabacum*) in China. Compared with the published genomes of eight strains which were isolated from other hosts and habitats, 794 specific genes and many rearrangements/inversion events were identified in the tobacco strain, demonstrating that this strain represents an important node within the *R. solanacearum* complex.**

*Ralstonia solanacearum* is a phytopathogenic bacterium with an unusually broad host range and a wide geographical distribution (1). It causes a wilt disease with deadly effects in many economically important crops like tomato, potato, tobacco, and banana (1, 5). Understanding the characteristics of *R. solanacearum* strains is helpful to determine the way to prevent or reduce potential economic losses. So far, eight genomes of *R. solanacearum* strains from three hosts (tomato, potato, and geranium) with a variety of habitats have been sequenced (6, 8, 14, 16, 19). Owing to the significant genomic variations of *R. solanacearum* strains derived from different regions and hosts, a strain (Y45) isolated from tobacco in Yunnan Province, China, was sequenced, annotated, and compared to the other eight strains. It is the first *R. solanacearum* strain isolated in East Asia.

Contigs and scaffolds were assembled and aligned using SOAPdenovo 1.05 (11) and SOAPaligner (10). Protein-coding and other genes were predicted by Glimmer3.0, rRNAmmer, tRNAscan, and Rfam (4, 7, 9, 17). Repetitive elements were identified by RepeatMasker (<http://www.repeatmasker.org>) and Tandem Repeat Finder (2). Phylogenetic and genomic analyses were carried out using MEGA4 based on *egl* and *hrpB* sequences (13, 18) and Mauve (15).

A total of 1,067.5 Mb clean data from two different insert size libraries (500 bp and 2 kb) were generated for strain Y45 using the Illumina HiSeq2000 paired-end strategy and assembled into 46 scaffold sequences (90% of the assembly falls into 10 scaffolds larger than 142 kb) with a length of approximately 5.73 Mb, representing 98.5% of the genome, whose size was estimated to be 5.82 Mb using the 17-mer approach (GSP; <http://gsizpred.sourceforge.net/>) (3, 12). Compared with the well-documented genome of strain GMI1000 (16), the 46 segments could be divided into three parts, including a chromosome segment (3,726,478 bp), a megaplasmid segment

(1,985,910 bp), and sequences which were not assigned to either of the two circular replicons, suggesting their uniqueness to this strain (17,900 bp). In all, 5,496 coding sequences, 53 tRNAs, and 5 rRNAs were predicted in the genome, respectively. Moreover, 14,472-bp transposable elements and 43,648-bp tandem repeats were predicted, comprising 0.25% and 0.76% of the genome, respectively. Of the 5,496 genes, 4,864 could produce a hit in at least one of the public protein databases (BLASTP *e* value,  $\leq 1e-5$ ); 632 genes, however, did not match any known proteins in the current public databases. In a comparative genome analysis with the eight sequenced *R. solanacearum* strains, 794 specific genes were identified within the genome of strain Y45, most of which are of unknown function or encode hypothetical proteins. Notably, genomic alignments revealed many rearrangements and inversions between strain Y45 (phylotype IB, according to our phylogenetic analysis) and the other strains, which occurred since Y45 separated from the other strains. Thus, the genome of strain Y45 represents a unique phylogenetic node and provides invaluable information for *R. solanacearum* genomic research.

**Nucleotide sequence accession numbers.** This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. AFWL00000000. The version described in this paper is the first version and is available under accession no. AFWL01000000.

We thank Weiliang Chen and Zhengyi Wang (Zhejiang University) for their help in DNA collection.

This project was supported by the China National Tobacco Corporation Yunnan Branch (2010YN01 and 2011YN04).

## REFERENCES

1. Álvarez, B., E. G. Biosca, and M. M. López. 2010. On the life of *Ralstonia solanacearum*, a destructive bacterial plant pathogen, p. 267–279. In A. Mendez-Vilas (ed.), *Current research, technology and education topics in applied microbiology and microbial biotechnology*, vol. 1. Formatex, Badajoz, Spain.
2. Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.
3. Chor, B., D. Horn, N. Goldman, Y. Levy, and T. Massingham. 2009. Genomic DNA k-mer spectra: models and modalities. *Genome Biol.* 10:R108.
4. Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg. 2007. Iden-

\* Corresponding author. Mailing address: Department of Agronomy & James D. Watson Institute of Genome Sciences, Zhejiang University, Hangzhou 310058, China. Phone: 86-571-88982730. Fax: 86-571-88982730. E-mail: fanlj@zju.edu.cn.

- tifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**:673–679.
5. **Elphinstone, J. G.** 2005. The current bacterial wilt situation: a global overview, p. 9–28. *In* C. Allen, P. Prior, and A. C. Hayward (ed.), *Bacterial wilt disease and the *Ralstonia solanacearum* species complex*. APS Press, St. Paul, MN.
  6. **Gabriel, D. W., et al.** 2006. Identification of open reading frames unique to a select agent: *Ralstonia solanacearum* race 3 biovar 2. *Mol. Plant Microbe Interact.* **19**:69–79.
  7. **Griffiths-Jones, S., et al.** 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**:D121–D124.
  8. **Guidot, A., et al.** 2009. Specific genes from the potato brown rot strains of *Ralstonia solanacearum* and their potential use for strain detection. *Phytopathology* **99**:1105–1112.
  9. **Lagesen, K., et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**:3100–3108.
  10. **Li, R., et al.** 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**:1966–1967.
  11. **Li, R., et al.** 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**:265–272.
  12. **Marçais, G., and C. Kingsford.** 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**:764–770.
  13. **Prior, P., and M. Fegan.** 2005. Recent development in the phylogeny and classification of *Ralstonia solanacearum*. *Acta Hort.* **695**:127–136.
  14. **Remenant, B., et al.** 2010. Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC Genomics* **11**:379.
  15. **Rissman, A. I., et al.** 2009. Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* **25**:2071–2073.
  16. **Salanoubat, M., et al.** 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**:497–502.
  17. **Schattner, P., A. N. Brooks, and T. M. Lowe.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**:W686–W689.
  18. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
  19. **Xu, J., et al.** 2011. Complete genome sequence of the plant pathogen *Ralstonia solanacearum* strain Po82. *J. Bacteriol.* **193**:4261–4262.